Hyperdynamics Importance Sampling

Cristian Sminchisescu, Bill Triggs

Hyperdynamics Importance Sampling

Cristian Sminchisescu and Bill Triggs

INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot, France.
{Cristian.Sminchisescu,Bill.Triggs}@inrialpes.fr
http://www.inrialpes.fr/movi/people/{Sminchisescu,Triggs}

Abstract

Sequential random sampling ('Markov Chain Monte-Carlo') is a popular strategy for many vision problems involving multimodal distributions over high-dimensional parameter spaces. It applies both to importance sampling *(where one wants to sample points according to their 'importance' for some calculation, but otherwise fairly) and to* global optimization *(where one wants to find good minima, or at least good starting points for local minimization, regardless of fairness). Unfortunately, most sequential samplers are very prone to becoming 'trapped' for long periods in unrepresentative local minima, which leads to biased or highly variable estimates. We present a general strategy for reducing MCMC trapping that generalizes Voter's 'hyperdynamic sampling' from computational chemistry. The local gradient and curvature of the input distribution are used to construct an adaptive importance sampler that focuses samples on low cost negative curvature regions likely to contain 'transition states' — codimension-1 saddle points representing 'mountain passes' connecting adjacent cost basins. This substantially accelerates inter-basin transition rates while still preserving correct relative transition probabilities. Experimental tests on the difficult problem of 3D articulated human pose estimation from monocular images show significantly enhanced minimum exploration.*

Keywords: *Hyperdynamics, Markov-chain Monte Carlo, importance sampling, global optimization, human tracking.*

1 Introduction

Many vision problems can be formulated either as global minimizations of highly non-convex cost functions with many minima, or as statistical inferences based on fair sampling or expectation-value integrals over highly multi-modal distributions. Importance sampling is a promising approach for such applications, particularly when combined with sequential ('Markov Chain Monte-Carlo'), layered or annealed samplers [8, 4, 5], optionally punctuated with bursts of local optimization [10, 3, 25]. Sampling methods are flexible, but they tend to be computationally expensive for a given level of accuracy. In particular, when used on multi-modal cost surfaces, current sequential samplers are very prone to becoming trapped for long periods in cost basins containing unrepresentative local minima. This 'trapping' or 'poor mixing' leads to biased or highly variable estimates whose character is at best quasi-local rather than global. Trapping times are typically exponential in a (large) scale parameter, so 'buying a faster computer' helps little. Current samplers are myopic mainly because they consider only the size of the integrand being evaluated or the lowness of the cost being optimized when judging 'importance'. *For efficient global estimates, it is also critically 'important' to include an*

To appear ECCV 2002

effective strategy for reducing trapping, e.g. *by explicitly devoting some fraction of the samples to moving between cost basins.*

This paper describes a method for reducing trapping by 'boosting' the dynamics of the sequential sampler. Our approach is based on Voter's 'hyperdynamics' [29, 30], which was originally developed in computational chemistry to accelerate the estimation of transition rates between different atomic arrangements in atom-level simulations of molecules and solids. There, the dynamics is basically a thermally-driven random walk of a point in the configuration space of the combined atomic coordinates, subject to an effective energy potential that models the combined inter-atomic interactions. The configuration-space potential is often highly multimodal, corresponding to different large-scale configurations of the molecule being simulated. Trapping is a significant problem, especially as the fine-scale dynamics must use quite short time-steps to ensure accurate physical modelling. Mixing times of 10^6–10^9 or more steps are common. In our target applications in vision the sampler need not satisfy such strict physical constraints, but trapping remains a key problem.

Hyperdynamics reduces trapping by boosting the number of samples that fall near 'transition states' — low lying saddle points that the system would typically pass through if it were moving thermally between adjacent energy basins. It does this by modifying the cost function, adding a term based on the gradient and curvature of the original potential that raises the cost near the cores of the local potential basins to reduce trapping there, while leaving the cost intact in regions where the original potential has the negative curvature eigenvalue and low gradient characteristic of transition neighborhoods. Hyperdynamics can be viewed as a generalized form of MCMC importance sampling whose importance measure considers the gradient and curvature as well as the values of the original cost function. The key point is not the specific form adopted for the potential, but rather the refined notion of 'importance': deliberately adding samples to speed mixing and hence reduce global bias ('finite sample effects'), even though the added samples are not directly 'important' for the calculation being performed.

Another general approach to multi-modal optimization is *annealing* — initially sampling with a reduced sensitivity to the underlying cost ('higher temperature'), then progressively increasing the sensitivity to focus samples on lower cost regions. Annealing has been used many times in vision and elsewhere[1], *e.g.* [18, 5], but although it works well in many applications, it has important limitations as a general method for reducing trapping. The main problem is that it samples indiscriminately within a certain energy band, regardless of whether the points sampled are likely to lead out of the basin towards another minimum, or whether they simply lead further up an ever-increasing potential wall. In many applications, and especially in high-dimensional or ill-conditioned ones, the cost surface has relatively narrow 'corridors' connecting adjacent basins, and it is important to steer the samples towards these using local information about how the cost appears to be changing. Hyperdynamics is a first attempt at doing this. In fact, these methods are complementary: it may be possible to speed up hyperdynamics by annealing its modified potential, but we will not investigate this here.

[1] Raising the temperature is often unacceptable in chemistry applications of hyperdynamics, as it may significantly change the problem. *E.g.*, the solid being simulated might melt...

1.1 What is a Good Multiple-Mode Sampling Function ?

'The curse of dimensionality' causes many difficulties in high-dimensional search. In stochastic methods, long sampling runs are often needed to hit the distribution's 'typical set' — the areas where most of the probability mass is concentrated. In sequential samplers this is due to the inherently local nature of the sampling process, which tends to become 'trapped' in individual modes, moving between them only very infrequently. More generally, choosing an importance sampling distribution is a compromise between tractable sampleability and efficient focusing of the sampling resources towards 'good places to look'.

There are at least three issues in the design of a good multi-modal sampler: (*i*) *Approximation accuracy*: in high dimensions, when the original distribution is complex and highly multi-modal (as is the case in vision), finding a good approximating function can be very difficult, thus limiting the applicability of the method. It is therefore appealing to look for ways of using a modified version of the original distribution, as for instance in annealing methods [18, 5]. (*ii*)*Trapping*: even when the approximation is locally accurate (*e.g.* by sampling the original distribution, thus avoiding any sample-weighting artifacts), most sampling procedures tend to get caught in the mode(s) closest to the starting point of sampling. Very long runs are needed to sample infrequent inter-mode transition events that lie far out in the tails of the modal distributions, but that can make a huge difference to the overall results. (*iii*)*Biased transition rates*: annealing changes not only the absolute inter-mode transition rates (thus reducing trapping), but also their relative sizes [27]. So there is no guarantee that the modes are visited with the correct relative probabilities implied by the dynamics on the original cost surface. This may seem irrelevant if the aim is simply to discover 'all good modes' or 'the best mode', but the levels of annealing needed to make difficult transitions frequent can very significantly increase the number of modes and the state space volume that are available to be visited, and thus cause the vast bulk of the samples to be wasted in fruitless regions[2]. This is especially important in applications like tracking, where only the neighboring modes that are separated from the current one by the lowest energy barriers need to be recovered.

To summarize, for complex high dimensional problems, finding good, sampleable approximating distributions is hard, so it is useful to look at sequential samplers based on distributions derived from the original one. There is a trade-off between sampling for local computational accuracy, which requires samples in 'important' regions, usually mode cores, and sampling for good mixing, which requires not only more frequent samples in the tails of the distribution, but also that these should be focused on regions likely to lead to inter-modal transitions. Defining such regions is delicate in practice, but it is clear that steering samples towards regions with low gradient and negative curvatures should increase the likelihood of finding transition states (saddle points with one negative curvature direction) relative to purely cost-based methods such as annealing.

[2] There is an analogy with the chemist's melting solid, liquids being regions of state space with huge numbers of small interconnected minima and saddles, while solids have fewer, or at least more clearly defined, minima. Also remember that state space volume increases very rapidly with sampling radius in high dimensions, so dense, distant sampling is simply infeasible.

1.2 Related Work

Now we briefly summarize some relevant work on high-dimensional search, especially in the domain of human modelling and estimation. Cham & Rehg [3] perform 2D tracking with scaled prismatic models. Their method combines a least squares intensity-based cost function, particle filtering with dynamical noise style sampling, and local optimization of a mixture of Gaussians state probability representation. Deutscher *et al* [5] track 3D body motion using a multi-camera silhouette-and-edge based likelihood function and annealed sampling within a temporal particle filtering framework. Their sampling procedure resembles one used by Neal [18], but Neal also includes an additional importance sampling correction designed to improve mixing. Sidenbladh *et al* [22] use an intensity based cost function and particle filtering with importance sampling based on a learned dynamical model to track a 3D model of a walking person in an image sequence. Choo & Fleet [4] combine particle filtering and hybrid Monte Carlo sampling to estimate 3D human motion, using a cost function based on joint re-projection error given input from motion capture data. Sminchisescu & Triggs [25] recover articulated 3D motion from monocular image sequences using an edge and intensity based cost function, with a combination of robust constraint-consistent local optimization and 'oversized' covariance scaled sampling to focus samples on probable low-cost regions.

Hyperdynamics uses stochastic dynamics with cost gradient based sampling as in [8, 17, 4], but 'boosts' the dynamics with a novel importance sampler constructed from the original probability surface using local gradient and curvature information. All of the annealing methods try to increase transition rates by sampling a modified distribution, but only the one given here specifically focuses samples on regions likely to contain transition states. There are also deterministic local-optimization-based methods designed to find transition states. See our companion paper [26] for references.

2 Sampling and Transition State Theory

2.1 Importance Sampling

Importance sampling works as follows. Suppose that we are interested in quantities depending on the distribution of some quantity $\mathbf{x}$, whose probability density is proportional to $f(\mathbf{x})$. Suppose that it is feasible to evaluate $f(\mathbf{x})$ pointwise, but that we are not able to sample directly from the distribution it defines, but only from an approximating distribution with density $f_b(\mathbf{x})$. We will base our estimates on a sample of N independent points, $\mathbf{x}_1, ..., \mathbf{x}_N$ drawn from $f_b(\mathbf{x})$. The expectation value of some quantity $V(\mathbf{x})$ with respect to $f(\mathbf{x})$ can then be estimated as $\bar{V} = \sum_{i=1}^{N} w_i\, V(\mathbf{x}_i) / \sum_{i=1}^{N} w_i$, where the **importance weighting** of $\mathbf{x}_i$ is $w_i = f(\mathbf{x}_i)/f_b(\mathbf{x}_i)$ (this assumes that $f_b(\mathbf{x}) \neq 0$ whenever $f(\mathbf{x}) \neq 0$). It can be proved that the importance sampled estimator converges to the mean value of V as N increases, but it is difficult to assess how reliable the estimate $\bar{V}$ is in practice. Two issues affect this accuracy: the variability of the importance weights due to deviations between $f(\mathbf{x})$ and $f_b(\mathbf{x})$, and statistical fluctuations caused by the improbability of sampling infrequent events in the tails of the distribution, especially if these are critical for estimating $\bar{V}$.

2.2 Stochastic Dynamics

Various methods are available for speeding up sampling. Here we use a stochastic dynamics method on the potential surface defined by our cost function (the negative log-likelihood of the state probability given the observations, $f(\mathbf{x}) = -\log p(\mathbf{x}|\cdot)$). Canonical samples from $f(\mathbf{x})$ can be obtained by simulating the phase space dynamics defined by the Hamiltonian function:

$$H(\mathbf{x}, \mathbf{p}) = f(\mathbf{x}) + K(\mathbf{p})$$

where $K(\mathbf{p}) = \mathbf{p}^\top \mathbf{p}/2$ is the kinetic energy, and $\mathbf{p}$ is the momentum variable. Averages of variables V over the canonical ensemble can be computed by using classical 2N-dimensional phase-space integrals:

$$\langle V \rangle \;=\; \frac{\iint V(\mathbf{x}, \mathbf{p}) e^{-\alpha f(\mathbf{x})} e^{-\alpha K(\mathbf{p})} \mathbf{dx dp}}{\iint e^{-\alpha f(\mathbf{x})} e^{-\alpha K(\mathbf{p})} \mathbf{dx dp}}$$

where $\alpha = 1/T$ is the temperature constant. Dynamics (and hence sampling) is done by locally integrating the Hamilton equations:

$$\frac{\mathbf{dx}}{dt} \;=\; \mathbf{p} \qquad \text{and} \qquad \frac{\mathbf{dp}}{dt} \;=\; -\frac{df(\mathbf{x})}{\mathbf{dx}}$$

using a Langevin Monte Carlo type integration/rejection scheme that is guaranteed to perform sampling from the canonical distribution over phase-space:

$$\mathbf{x}_{i+1} \;=\; \mathbf{x}_i - \frac{\Delta t_{sd}^2}{2} \frac{df(\mathbf{x})}{\mathbf{dx}} + \Delta t_{sd} \mathbf{n}_i \tag{1}$$

where $\mathbf{n}_i$ is a vector of independently chosen Gaussian variables with zero mean and unit variance, and Δt_{sd} is the stochastic dynamics integration step. Compared to so called 'hybrid' methods, the Langevin method can be used with a larger step size and this is advantageous for our problem, where the step calculations are relatively expensive (see [17] and its references for a more complete discussion of the relative advantages of hybrid and Langevin Monte Carlo methods)[3]. For physical dynamics t represents the physical time, while for statistical calculations it simply represents the number of steps performed since the start of the simulation. The simulation time is used in §3 below to estimate the acceleration of infrequent events produced by the proposed biased potential.

2.3 Transition State Theory

Continuing the statistical mechanics analogy begun in the previous section, the behavior of the physical system can be characterized by long periods of 'vibration' within

[3] Note that the momenta are only represented implicitly in the Langevin formulation: there is no need to update their values after each leapfrog step as they are immediately replaced by new ones drawn from the canonical distribution at the start of each iteration. If approximate cost Hessian information is also available, the gradient in (1) can be projected onto the Hessian eigen-basis and its components weighted by the local eigen-curvatures to give an effective 'Newton-like' step. We use such steps near saddle points, where the hyperdynamic bias potential is essentially zero, to avoid the inefficiencies of random walk behavior there.

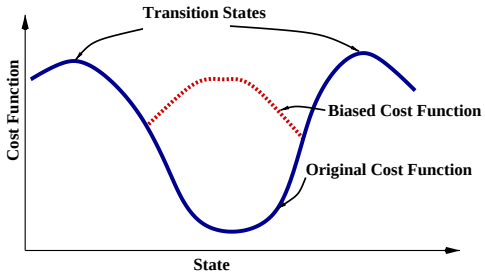

Fig. 1. The original cost function and the bias added for hyperdynamics.

one 'state' (energy basin), followed by infrequent transitions to other states via saddle points. In the 'transition state theory' (TST) approximation, the transition rates between states are computed using the sample flux through the *dividing surface* separating them. For a given state S, this is the $N-1$ dimensional surface separating the state S from its neighbors. The rate of escape from state S is:

$$k^{tst}_{S\rightarrow} \;=\; \langle\, |\nu_S|\, \delta_S(\mathbf{x}) \,\rangle_S$$

where $\delta_s(\mathbf{x})$ is a Dirac delta function positioned on the dividing surface of S and ν_s is the velocity normal to this surface. Crossings of the dividing surface correspond to true state change events, and we assume that the system loses all memory of this transition before the next event.

3 Accelerating Transition State Sampling

In the above formalism, the TST rate can be evaluated as follows:

$$k^{tst}_{S\rightarrow} \;=\; \frac{\iint\, |\nu_S|\, \delta_S(\mathbf{x})\, e^{-\alpha f(\mathbf{x})}\, e^{-\alpha K(\mathbf{p})}\, \mathbf{dx}\, \mathbf{dp}}{\iint\, e^{-\alpha f(\mathbf{x})}\, e^{-\alpha K(\mathbf{p})}\, \mathbf{dx}\, \mathbf{dp}}$$

Now consider adding a positive bias or boost cost $f_b(\mathbf{x})$ (with a corresponding 'biased' state S_b) to the original cost $f(\mathbf{x})$, with the further property that $f_b(\mathbf{x}) = 0$ whenever $\delta_S(\mathbf{x}) \neq 0$, *i.e.* the potential is unchanged in the transition state regions. The TST rate becomes:

$$k^{tst}_{S\rightarrow} \;=\; \frac{\iint\, |\nu_S|\, \delta_S(\mathbf{x})\, e^{-\alpha[f(\mathbf{x})+f_b(x)]}\, e^{\alpha f_b(\mathbf{x})}\, e^{-\alpha K(\mathbf{p})}\, \mathbf{dx}\, \mathbf{dp}}{\iint\, e^{-\alpha f(\mathbf{x})}\, e^{-\alpha K(\mathbf{p})}\, \mathbf{dx}\, \mathbf{dp}} \tag{2}$$

$$=\; \frac{\langle\, |\nu_S|\, \delta_S(\mathbf{x})\, e^{\alpha f_b(\mathbf{x})} \,\rangle_{S_b}}{\langle\, e^{\alpha f_b(\mathbf{x})} \,\rangle_{S_b}} \;=\; \frac{\langle\, |\nu_S|\, \delta_S(\mathbf{x}) \,\rangle_{S_b}}{\langle\, e^{\alpha f_b(\mathbf{x})} \,\rangle_{S_b}} \tag{3}$$

The boost term increases every escape rate from state S as the cost well is made shallower, but it leaves the *ratios* of escape rates from S, S_b to other states S_1, S_2 invariant:

$$\frac{k^{tst}_{S\rightarrow S_1}}{k^{tst}_{S\rightarrow S_2}} \;=\; \frac{k^{tst}_{S_b\rightarrow S_1}}{k^{tst}_{S_b\rightarrow S_2}}$$

This holds because all escape rates from S all have the partition function of S as denominator, and replacing this with the partition function of S_b leaves their ratios unchanged.

Concretely, suppose that during N_t steps of classical dynamics simulation on the biased cost surface, we encounter N_e escape attempts over the dividing surface. For the computation, let us also assume that the simulation is artificially confined to the basin of state S by reflecting boundaries. (This does not happen in real simulations: it is used here only to estimate the 'biased boost time'). The TST escape rate from state S can be estimated simply as the ratio of the number of escape attempts to the total trajectory length: $k_S^{tst} = N_e/(N_t \Delta t_{sd})$. Consequently, the mean escape time (inverse transition rate) from state S can be estimated from (2) as:

$$\tau_{esc}^S \;=\; \frac{1}{k_{S\rightarrow}^{tst}} \;=\; \frac{\langle\, e^{\alpha f_b(\mathbf{x})} \,\rangle_{S_b}}{\langle\, |\nu_S|\, \delta_S(\mathbf{x}) \,\rangle_{S_b}} \;=\; \frac{\frac{1}{N_t}\sum_{i=1}^{N_t} e^{\alpha f_b(\mathbf{x}_i)}}{N_e/(N_t\, \Delta t_{sd})} \;=\; \frac{1}{N_e}\sum_{i=1}^{N_t} \Delta t_{sd}\, e^{\alpha f_b(\mathbf{x}_i)}$$

The effective simulation time boost achieved in step i thus becomes simply:

$$\Delta t_{b_i} \;=\; \Delta t_{sd} e^{\alpha f_b(\mathbf{x}_i)} \tag{4}$$

The dynamical evolution of the system from state to state is still correct, but it works in a distorted time scale that depends exponentially on the bias potential. As the system passes through regions with high f_b, its equivalent time Δt_b increases rapidly as it would originally have tended to linger in these regions (or more precisely to return to them often on the average) owing to their low original cost. Conversely, in zones with small f_b the equivalent time progress at the standard stochastic dynamics rate. Of course, in reality the simulation's integration time step and hence its sampling coarseness are the same as they were in the unboosted simulation. The boosting time (4) just gives an intuition for how much time an unaccelerated sampler would probably have wasted making 'uninteresting' samples near the cost minimum. But that is largely the point: the wastage factors are astronomical in practice — unboosted samplers can not escape from local minima.

4 The Biased Cost

The main requirements on the bias potential are that it should be zero on all dividing surfaces, that it should not introduce new sub-wells with escape times comparable to the main escape time from the original cost well, and that its definition should not require prior knowledge of the cost wells or saddle points (if we knew these we could avoid trapping much more efficiently by including explicit well-jumping samples). For sampling, the most 'important' regions of the cost surface are minima, where the Hessian matrix $\mathbf{H}$ has strictly positive eigenvalues, and transition states, where it has exactly one negative eigenvalue $e_1 < 0$. The gradient vector vanishes in both cases. The rigorous definition of the TST boundary is necessarily global[4], but locally near a transition state the boundary contains the state itself and adjacent points where the Hessian has a negative eigenvalue and vanishing gradient component along the corresponding eigenvector:

$$g_{p1} \;=\; \mathbf{V}_1^\top \mathbf{g} \;=\; 0 \qquad \text{and} \qquad e_1 < 0 \tag{5}$$

[4] The basin of state S can be defined as the set of configurations from which gradient descent minimization leads to the minimum S. This basin is surrounded by an $(n-1)$-D hypersurface, outside of which local descent leads to states other than S.

where $\mathbf{g}$ is the gradient vector and $\mathbf{V}_1$ is the first Hessian eigenvector. Voter [29, 30] therefore advocates the following bias cost for hyperdynamics:

$$f_b \;=\; \frac{h_b}{2}\left[1 + \frac{e_1}{\sqrt{e_1^2 + g_{p1}^2/d^2}}\right] \tag{6}$$

where h_b is a constant controlling the strength of the bias and d is a length scale (*e.g.* an estimate of the typical nearest-neighbour distance between minima, if this is available). Note that Voter's f_b has all of the properties required in §3. In particular, it is zero on the dividing surface, as can be seen from (5) and (6).

Increasing h_b increases the bias and hence the nominal boosting. In principle it is even permissible to raise the cost of a minimum above the level of its surrounding transition states. However, there is a risk that doing so will entirely block the sampling pathways through and around the minimum, thus causing the system to become trapped in a newly created well at one end of the old one. Hence, it is usually safer to select a more moderate boosting.

One difficulty with Voter's potential (6) is that direct differentiation of it for gradient-based dynamics requires third order derivatives of $f(\mathbf{x})$. However an inexpensive numerical estimation method based on first order derivatives was proposed in [30]. For completeness we summarize this in the appendix. These calculations are more complex than those needed for standard gradient based stochastic simulation, but we will see that the bias provides a degree of acceleration that often pays-off in practice.

5 Human Domain Modelling

This section briefly describes the humanoid visual tracking models used in our hyperdynamic boosting experiments. For more details see [24, 25].

Representation: Our body models contain kinematic 'skeletons' of articulated joints controlled by angular joint parameters, covered by 'flesh' built from superquadric ellipsoids with additional global deformations [1]. A typical model has about 30-35 joint parameters $\mathbf{x}_a$; 8 internal proportion parameters $\mathbf{x}_i$ encoding the positions of the hip, clavicle and skull tip joints; and 9 deformable shape parameters for each body part, gathered into a vector $\mathbf{x}_d$. The complete model is thus encoded as a single large parameter vector $\mathbf{x} = (\mathbf{x}_a, \mathbf{x}_d, \mathbf{x}_i)$. During tracking or static pose estimation we usually estimate only joint parameters.

The model is used as follows. Superquadric surfaces are discretized into meshes parameterized by angular coordinates in a 2D topological domain. Mesh nodes $\mathbf{u}_i$ are transformed into 3D points $\mathbf{p}_i(\mathbf{x})$, then into predicted image points $\mathbf{r}_i(\mathbf{x})$ using composite nonlinear transformations $\mathbf{r}_i(\mathbf{x}) = P(\mathbf{p}_i(\mathbf{x})) = P(A(\mathbf{x}_a, \mathbf{x}_i, D(\mathbf{x}_d, \mathbf{u}_i)))$, where D represents a sequence of parametric deformations that construct the corresponding part in its own reference frame, A represents a chain of rigid transformations that map it through the kinematic chain to its 3D position, and P represents perspective image projection. During model estimation, prediction-to-image matching cost metrics are evaluated between each predicted model feature $\mathbf{r}_i$ and nearby associated image features $\bar{\mathbf{r}}_i$, and the results are summed over all features to produce the image contribution to the

overall parameter space cost function. The cost is thus a robust function of the prediction errors $\boldsymbol{\Delta}\mathbf{r}_i(\mathbf{x}) = \bar{\mathbf{r}}_i - \mathbf{r}_i(\mathbf{x})$. The cost gradient $\mathbf{g}_i(\mathbf{x})$ and Hessian $\mathbf{H}_i(\mathbf{x})$ are also computed and assembled over all observations.

Estimation: We aim for a probabilistic interpretation and optimal estimates of the model parameters by maximizing the total probability according to Bayes rule:

$$p(\mathbf{x}|\bar{\mathbf{r}}) \;\propto\; p(\bar{\mathbf{r}}|\mathbf{x})\, p(\mathbf{x}) \;=\; \exp\left(-\textstyle\int e(\bar{\mathbf{r}}_i|\mathbf{x})\, di\right)\; p(\mathbf{x}) \tag{7}$$

where $e(\bar{\mathbf{r}}_i|\mathbf{x})$ is the cost density associated with observation i, the integral is over all observations, and $p(\mathbf{x})$ is the prior on the model parameters. Discretizing the continuous problem, our MAP approach minimizes the negative log-likelihood for the total posterior probability:

$$f(\mathbf{x}) \;=\; -\log p(\bar{\mathbf{r}}|\mathbf{x}) - \log p(\mathbf{x}) \;=\; f_l(\mathbf{x}) + f_p(\mathbf{x}) \tag{8}$$

Observation Likelihood: In the below experiments we actually only used a very simple Gaussian likelihood based on given model-to-image joint correspondences. The negative log-likelihood for the observations is just the sum of squared model joint reprojection errors. Our full tracking system uses this cost function only for initialization, but it still provides an interesting (and difficult to handle) degree of multimodality owing to the kinematic complexity of the human model and the large number of parameters that are unobservable in a singular monocular image. In practice we find that globalizing the search is at least as important for initialization as for tracking, and this cost function is significantly cheaper to evaluate than our full image based one, allowing more extensive sampling experiments.

Priors and Constraints: Both hard and soft priors are accommodated in our framework. They include anthropometric priors on model proportions, parameter stabilizers for hard to estimate but useful modelling parameters, terms for collision avoidance between body parts, and joint angle limits. During estimation, the values, gradients and Hessians of the priors are evaluated and added to the contributions from the observations.

6 Experiments and Results

In this section we illustrate the hyperdynamics method on a toy problem involving a two-dimensional multi-modal cost surface, and on the problem of initial pose estimation for an articulated 3D human model based on given joint-to-image correspondences. In both cases we compare the method with standard stochastic dynamics on the original cost surface. The parameters of the two methods (temperature, integration step, number of simulation steps, *etc*.) are identical, except that hyperdynamics requires values for the two additional parameters h_b and d that control the properties of the bias potential (6).

6.1 The Müller Cost Surface

Müller's Potential (fig. 2, left) is a simple 2D analytic cost function with three local minima M_1, M_2, M_3, and two saddle points S_1, S_2, which is often used in the chemistry

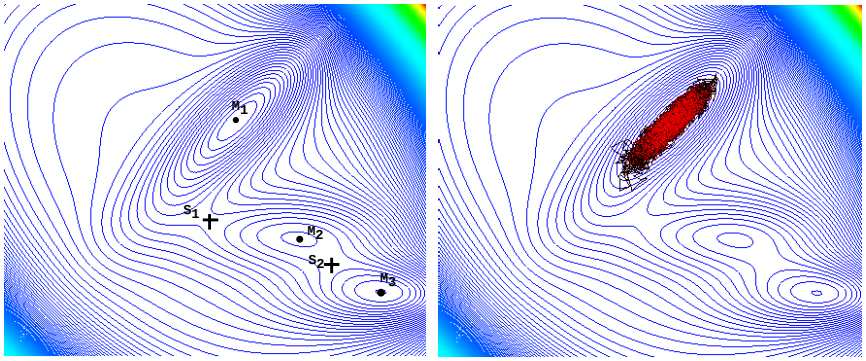

Fig. 2. The Müller Potential (left) and a standard stochastic dynamics gradient sampling simulation (right) that gets trapped in the basin of the starting minimum.

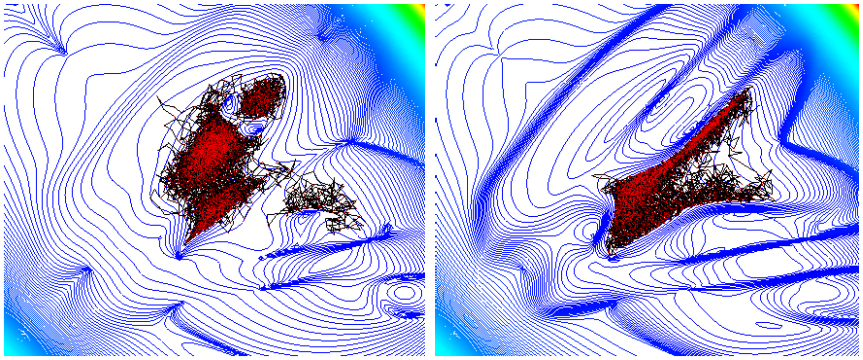

Fig. 3. Hyperdynamic sampling with $h_b = 150, d = 0.1$ and $h_b = 200, d = 0.5$.

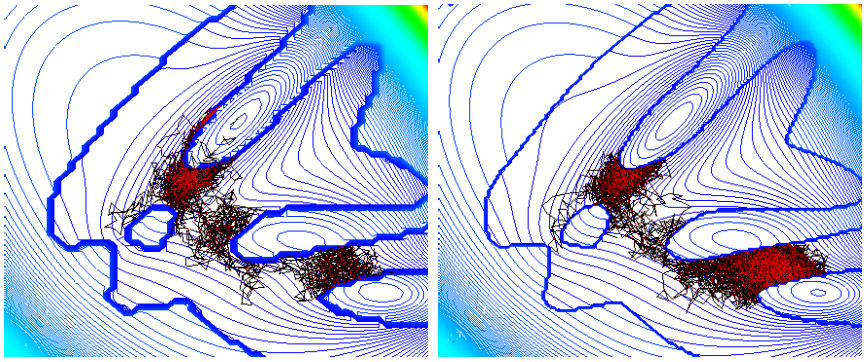

Fig. 4. Hyperdynamic sampling with $h_b = 300, d = 10$ and $h_b = 400, d = 100$.

literature to illustrate transition state search methods[5]. The inter-minimum distance is

[5] It has the form $V(x,y) = \sum_{i=1}^{4} A_i\, e^{a_i(x-x_i)^2+b_i(x-x_i)(y-y_i)+c_i(y-y_i)^2}$ where $\mathbf{A} = (-200,-100,-170,15)$, $\mathbf{a} = (-1,-1,-6.5,0.7)$, $\mathbf{b} = (0,0,11,0.6)$, $\mathbf{c} = (-10,-10,-6.5,0.7)$, $\mathbf{x} = (1,0,-0.5,-1)$, $\mathbf{y} = (0,0.5,1.5,1)$.

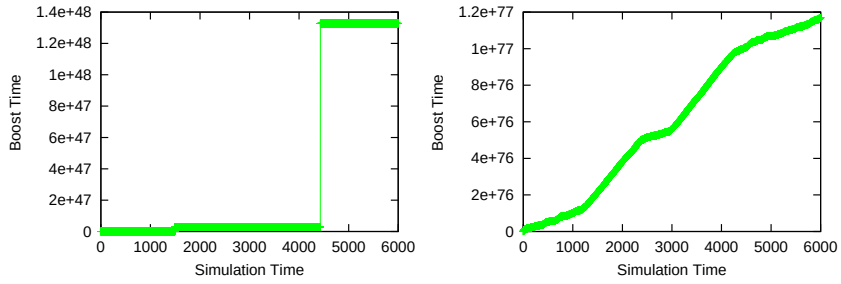

Fig. 5. Effective boost times for mild (left) and more aggressive (right) bias potentials.

of order 1 length unit, and the transition states are around 100–150 energy units above the lowest minimum.

Fig. 2(right) shows the result of standard stochastic dynamic sampling on the original cost surface. Despite 6000 simulation steps at a reasonable step size $\Delta t_{sd} = 0.01$, only the basin of the starting minimum is sampled extensively, and no successful escape has yet taken place. Fig. 3 shows two hyperdynamics runs with parameters set for moderate boosting. Note the reduced emphasis on sampling in the core of the minimum — in fact the minimum is replaced by a set of higher energy ones — and the fact that the runs escape the initial basin. In the right hand plot there is a clear focusing of samples in the region corresponding to the saddle point linking the two adjacent minima M_1 and M_2. Finally, fig. 4 shows results for more aggressive bias potentials that cause the basins of all three minima to be visited, with strong focusing of samples on the inter-minimum transition regions. The bias here turns the lowest positive curvature region of the initial minimum into a local maximum.

The plots also show that the Voter potential is somewhat 'untidy', with complicated local steps and ridges. Near the hypersurfaces where the first Hessian eigenvalue e_1 passes down through zero, the bias jumps from h_b to 0 with an abruptness that increases as the length scale d increases (sic) or the gradient projection g_{p1} decreases, owing to the $e_1/\sqrt{e_1^2 + g_{p1}^2/d^2}$ term in (6). A small d makes these $e_1 = 0$ transitions smoother, but increases the suddenness of ridges in the potential that occur on hypersurfaces where g_{1p} passes through zero.

Fig. 5 plots the simulation boosting time for two bias potentials. The left plot has a milder potential that simply encourages exploration of saddle points, while the right plot has a more aggressive one that is able to explore and jump between individual modes more rapidly. (Note the very large and very different sizes of the boosting time scales in these plots).

6.2 Monocular 3D Pose Estimation

Now we explore the potential of the hyperdynamics method for monocular 3D human pose estimation under model to image joint correspondences. This problem is well adapted to illustrating the algorithm, as its cost surface is highly multimodal. Of the 32

kinematic model d.o.f., about 10 are subject to 'reflective' kinematic ambiguities (forwards *vs*. backwards slant in depth), which potentially creates around $2^{10} = 1024$ local minima in the cost surface [13], although some of these are not physically feasible and are automatically pruned during the simulation (see below). Indeed, we find that it is very difficult to ensure initialization to the 'correct' pose with this kind of data.

The simulation enforces joint limit constraints using reflective boundary conditions, *i.e*. by reversing the sign of the particle's normal momentum when it hits a joint limit. We found that this gives an improved sampling acceptance rate compared to simply projecting the proposed configuration back into the constraint surface, as the latter leads to cascades of rejected moves until the momentum direction gradually swings around.

We ran the simulation for 8000 steps with $\Delta t_{sd} = 0.01$, both on the original cost surface (fig. 8) and on the boosted one (fig. 6). It is easy to see that the original sampler gets trapped in the starting mode, and wastes all of its samples exploring it repeatedly. Conversely, the boosted hyperdynamics method escapes from the starting mode relatively quickly, and subsequently explores many of the minima resulting from the depth reflection ambiguities.

Fig. 7 plots the estimated boosting times for two different bias potentials, $h_b = 200, d = 2$, and $h_b = 400, d = 20$. The computed mean state variance of the original estimator was 4.10^{-6}, compared to 7.10^{-6} for the boosted one.

7 Conclusions and Future Work

We underlined the fact that for global investigation of strongly multimodal high dimensional cost functions, importance samplers need to devote some of their samples to reducing trapping in local minima, rather than focusing only on performing their target computation. With this in mind, we presented an MCMC sampler designed to accelerate the exploration of different minima, based on the 'hyperdynamics' method from computational chemistry. It uses local cost gradients and curvatures to construct a modified cost function that focuses samples towards regions with low gradient and at least one negative curvature, which are likely to contain the transition states (low cost saddle points with one negative curvature direction) of the original cost. Our experimental results demonstrate that the method significantly improves inter-minimum exploration behaviour in the problem of monocular articulated 3D human pose estimation.

Our future work will focus on deriving alternative, computationally more efficient biased sampling distributions.

Acknowledgements This work was supported by an EIFFEL doctoral grant and European Union FET-Open project VIBES. We would like to thank Alexandru Telea for implementation discussions.

Appendix: Estimating the Gradient of Voter's Potential

Direct calculation of the gradient of Voter's potential (6) requires third order derivatives of $f(\mathbf{x})$, but an inexpensive numerical estimation method based on first order derivatives

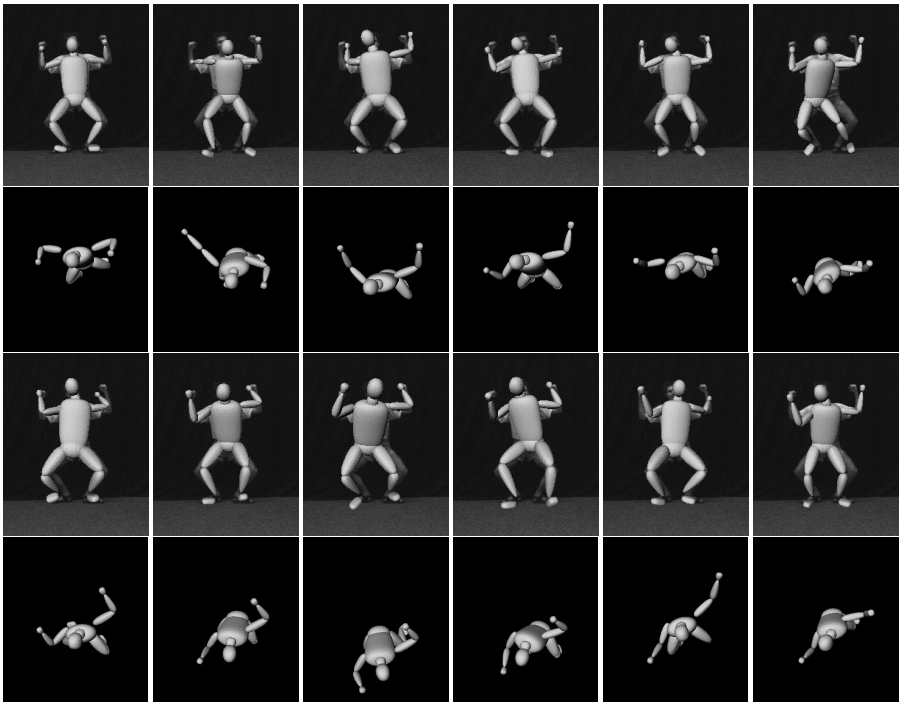

Fig. 6. Human poses sampled using hyperdynamics on a cost surface based on given model-to-image joint correspondences, seen from the camera viewpoint and from above. Hyperdynamics finds a variety of different poses including well separated reflective ambiguities (which, as expected, all look similar from the camera viewpoint). In contrast, standard stochastic dynamics (on the same underlying cost surface with identical parameters) essentially remains trapped in the original starting mode even after 8000 simulation steps (fig. 8).

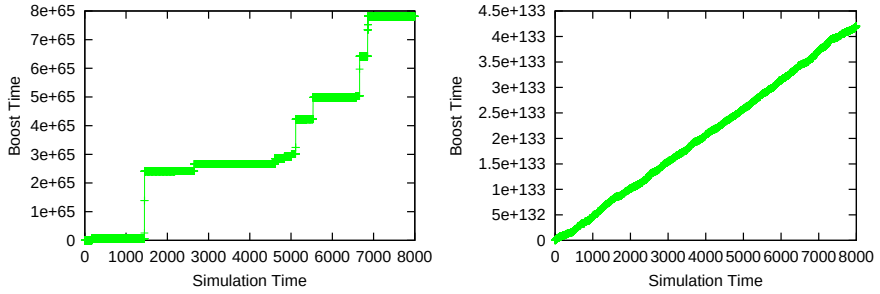

Fig. 7. Boosting times for human pose experiments, with mild (left) and strong (right) bias.

was proposed in [30]. An eigenvalue can be computed by numerical approximation along it's corresponding eigenvector direction $\mathbf{s}$:

$$e(\mathbf{s}) \;=\; [f(\mathbf{x}+\eta\mathbf{s}) + f(\mathbf{x}-\eta\mathbf{s}) - 2f(\mathbf{x})]/\eta^2 \tag{9}$$

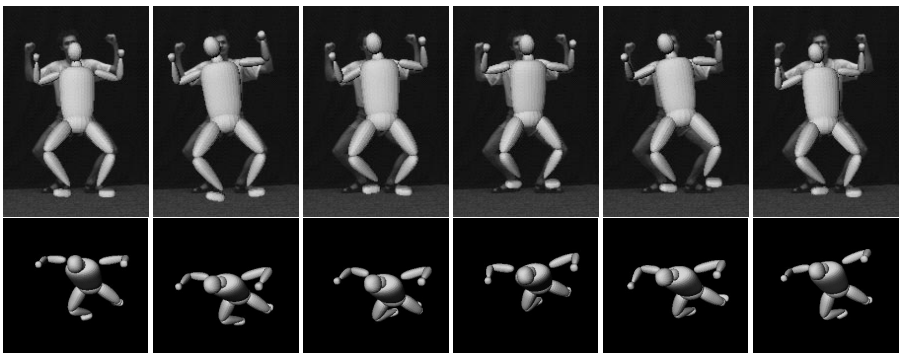

Fig. 8. Stochastic dynamics on the original cost surface leads to "trapping" in the starting mode.

The eigenvector direction can be estimated numerically using any gradient descent method, based on a random initialization $\mathbf{s}$ or on the one from the previous dynamics step, using:

$$\frac{de}{d\mathbf{s}} \;=\; [\mathbf{g}(\mathbf{x}+\eta\mathbf{s}) - \mathbf{g}(\mathbf{x}-\eta\mathbf{s})]/\eta \tag{10}$$

The lowest eigenvector obtained from the minimization (10) is then used to compute the corresponding eigenvalue via (9). The procedure can be repeated for higher eigenvalue-eigenvector pairs by maintaining orthogonality with previous directions. The derivative of the projected gradient g_{1p} can then be obtained by applying the minimization to the matrices $\mathbf{H} + \lambda\,\mathbf{g}\,\mathbf{g}^\top$ and $\mathbf{H} - \lambda\,\mathbf{g}\,\mathbf{g}^\top$. One thus minimizes:

$$\frac{de_i}{\mathbf{dx}} \;=\; \left\{[\mathbf{g}(\mathbf{x}+\eta\mathbf{s}) + \mathbf{g}(\mathbf{x}-\eta\mathbf{s}) - 2\mathbf{g}(\mathbf{x})]/\eta^2\right\}_{\mathbf{s}=\mathbf{s}_i}$$

where:

$$e_{\pm\lambda} \;=\; e(\mathbf{s}) \pm \lambda\Big[\frac{f(\mathbf{x}+\eta\mathbf{s}) - f(\mathbf{x}-\eta\mathbf{s})}{2\eta}\Big]^2$$

A good approximation to g_{p1} can be obtained from [30]:

$$g_{p1} \;=\; \frac{1}{2\lambda}(e_{+\lambda} - e_{-\lambda}), \qquad \text{and} \qquad \frac{dg_{p1}}{\mathbf{dx}} \;=\; \frac{1}{2\lambda}\left(\frac{de_{+\lambda}}{\mathbf{dx}} - \frac{de_{-\lambda}}{\mathbf{dx}}\right)$$

References

[1] A. Barr. Global and Local Deformations of Solid Primitives. *Computer Graphics*, 18:21–30, 1984.

[2] M. Black and A. Rangarajan. On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision. *IJCV*, 19(1):57–92, July 1996.

[3] T. Cham and J. Rehg. A Multiple Hypothesis Approach to Figure Tracking. In *CVPR*, volume 2, pages 239–245, 1999.

[4] K. Choo and D. Fleet. People Tracking Using Hybrid Monte Carlo Filtering. In *ICCV*, 2001.

[5] J. Deutscher, A. Blake, and I. Reid. Articulated Body Motion Capture by Annealed Particle Filtering. In *CVPR*, 2000.
[6] J. Deutscher, B. North, B. Bascle, and A. Blake. Tracking through Singularities and Discontinuities by Random Sampling. In *ICCV*, pages 1144–1149, 1999.
[7] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.
[8] D. Forsyth, J. Haddon, and S. Ioffe. The Joy of Sampling. *IJCV*, 41:109–134, 2001.
[9] D. Gavrila and L. Davis. 3-D Model Based Tracking of Humans in Action:A Multiview Approach. In *CVPR*, pages 73–80, 1996.
[10] T. Heap and D. Hogg. Wormholes in Shape Space: Tracking Through Discontinuities Changes in Shape. In *ICCV*, pages 334–349, 1998.
[11] N. Howe, M. Leventon, and W. Freeman. Bayesian Reconstruction of 3D Human Motion from Single-Camera Video. *ANIPS*, 1999.
[12] O. King and D. Forsyth. How does CONDENSATION Behave with a Finite Number of Samples? In *ECCV*, pages 695–709, 2000.
[13] H. J. Lee and Z. Chen. Determination of 3D Human Body Postures from a Single View. *CVGIP*, 30:148–168, 1985.
[14] J. MacCormick and M. Isard. Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracker. In *ECCV*, volume 2, pages 3–19, 2000.
[15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.
[16] D. Morris and J. Rehg. Singularity Analysis for Articulated Object Tracking. In *CVPR*, pages 289–296, 1998.
[17] R. Neal. Probabilistic Inference Using Markov Chain Monte Carlo. Technical Report CRG-TR-93-1, University of Toronto, 1993.
[18] R. M. Neal. Annealed Importance Sampling. *Statistics and Computing*, 11:125–139, 2001.
[19] R. Plankers and P. Fua. Articulated Soft Objects for Video-Based Body Modeling. In *ICCV*, pages 394–401, 2001.
[20] R. Rosales and S. Sclaroff. Inferring Body Pose without Tracking Body Parts. In *CVPR*, pages 721–727, 2000.
[21] E. M. Sevick, A. T. Bell, and D. N. Theodorou. A Chain of States Method for Investigating Infrequent Event Processes Occuring in Multistate, Multidimensional Systems. *J. Chem. Phys.*, 98(4), 1993.
[22] H. Sidenbladh, M. Black, and D. Fleet. Stochastic Tracking of 3D Human Figures Using 2D Image Motion. In *ECCV*, 2000.
[23] C. Sminchisescu. Consistency and Coupling in Human Model Likelihoods. In *CFGR*, 2002.
[24] C. Sminchisescu and B. Triggs. A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking. Technical Report RR-4208, INRIA, 2001.
[25] C. Sminchisescu and B. Triggs. Covariance-Scaled Sampling for Monocular 3D Body Tracking. In *CVPR*, 2001.
[26] C. Sminchisescu and B. Triggs. Building Roadmaps of Local Minima of Visual Models. In *ECCV*, 2002.
[27] M. R. Sorensen and A. F. Voter. Temperature-Accelerated Dynamics for Simulation of Infrequent Events. *J. Chem. Phys.*, 112(21):9599–9606, 2000.
[28] G. H. Vineyard. Frequency factors and Isotope Effects in Solid State Rate Processes. *J. Phys. Chem. Solids*, 3:121–127, 1957.
[29] A. F. Voter. A Method for Accelerating the Molecular Dynamics Simulation of Infrequent Events. *J. Chem. Phys.*, 106(11):4665–4677, 1997.
[30] A. F. Voter. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Physical Review Letters*, 78(20):3908–3911, 1997.